Available online at www.elixirpublishers.com (Elixir International Journal)

Computer Science and Engineering

Elixir Comp. Sci. & Engg. 63 (2013) 18149-18156

ISSN: 2229-712X

Human action recognition to understand hand signals for traffic surveillance

R.Sathya* and M. Kalaiselvi Geetha

Department of Computer Science and Engineering, Annamalai University.

ARTICLE INFO

Article history:

Received: 30 August 2013;

Received in revised form:

29 September 2013;

Accepted: 1 October 2013;

Keywords

Computer vision,
Hand action,
Feature extraction,
Activity detection,
Activity recognition.

ABSTRACT

Gesture Recognition plays a vital role in computer vision. The purpose of this survey is to provide a detailed overview and categories of current issues and trends. The recognition of human hand gesture movement can be performed at various levels of abstraction. Many applications and algorithms were discussed here with the explanation of system recognition framework. General overview of an action and its various applications were discussed in this paper. Most of the recognition system uses the data sets like KTH, Weizmann. Some other data sets were used by the action recognition system. In this paper, various approaches for image representation, feature extraction, activity detection and action recognition were also discussed.

Introduction

Hand movement recognition in video is an important application area of research in computer vision. It is the interpretation of movement of the hands, arms or body parts that gives a semantic meaning. Hand movement can be more useful particularly at a long distance, where speech information is not available. It has a wide range of applications such as interacting with computers through hand movements, machine vision, understanding signs of traffic personal for traffic surveillance and automated vehicles etc.,

Hand movement recognition can be viewed at two levels: hand posture and hand gesture recognition. Hand posture is the static pose of finger and palm, for example, thumbs up, pointing etc. Where as a hand gesture is the dynamic movement, that involves transformation of hand position and orientation such as calling, stopping, waving etc., A human can specify these indications for controlling robots by their gesture. Further, hand movement recognition differs from gait recognition, where the latter focuses on the individual walking styles that can be utilized as biometric information.

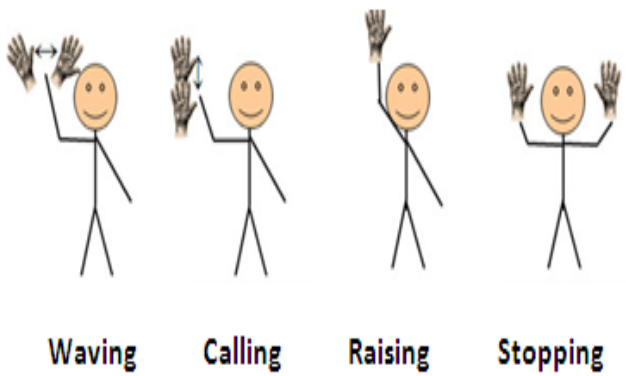

Fig. 1. Human Robot Interaction [1]

An action recognition system requires the ability to distinguish between target gestures from user's common actions such as scratching their head, clasping their hands above their head, scratching their body with raised hands and other gestures [1]. Moreover an action is done normally with a number of successive actions, which gives an interpretation of the action carried out. Recent literature focus in the area of vision based analysis of human poses and actions from video sequence.

General Architecture

Human action recognition approaches are normally discussed as vision based approaches and data glove based approaches in the literature. The focus of this paper is limited to vision based human action recognition approaches. This paper addresses the issues at different levels as seen in Fig. 2. Moreover this paper point out the challenges involved and discusses the future directions.

The input video is decomposed into a set of features tacking individual frames into account. The hands are isolated from their body parts as well as other background objects. Appearance based approach method uses image features to model the visual appearance of the hand and compare these parameters with the extracted image features from the video input.

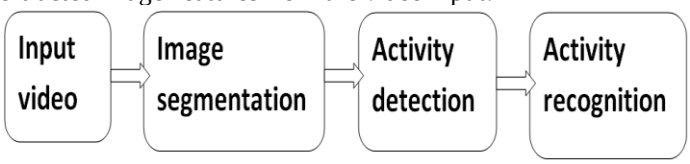

Fig. 2. General architecture

In this general architecture, many papers have been focused with research detailing several techniques for the segmentation, activity detection and activity recognition. The paper is structured as follows. Immediately below, detailed about common data sets. In section 5 explain the image segmentation. In section 6 detailed the form of human activity detection. The activity recognition techniques are presented in section 7. In section 8 discuss about some challenges. In section 9 discuses about many application areas. Feature direction focused on section 10. Section 11 concludes the paper.

Tele:
E-mail addresses: rsathyamephd@gmail.com

Common Datasets

KTH Dataset

The KTH human action dataset is a mostly exploited while the very challenging dataset. It contains six actions such as walking, jogging, boxing, hand waving and hand clapping. It contains six hundred video sequences. Each video has only one action performed by twenty five different actors. Four different scenarios are used: Outdoors, outdoors with scale variations, outdoors with different cloths, and indoors. The backgrounds are relatively static (no background noise). Apart from the zooming scenarios, there is only slight camera movement.

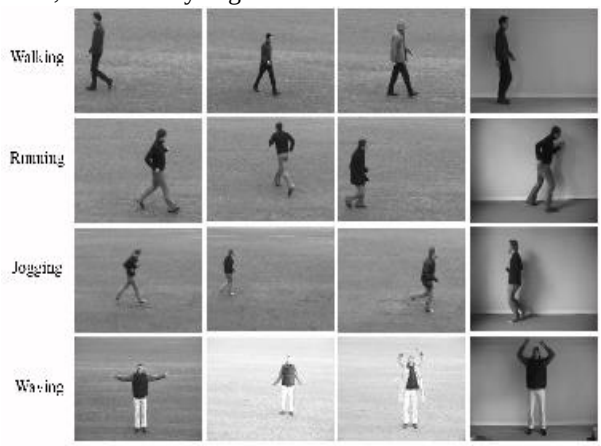

Fig. 3. Example frames for KTH Dataset

Weizmann Dataset

Weizmann human action Datasets contain ten actions such as walking, running, jumping-jack, jumping-forward-on-two-legs, jumping-in-place-on-two-legs, waving-two-hands, waving-one hand, bending, galloping-sideways, performed by nine different actors. The view point is static. Two separate sets are available. One set shows walking movement viewed from different angles. The second set shows front to parallel walking actions with slight variations (carrying objects, different clothing, and different styles). The background is static and foreground silhouettes are included in the dataset.

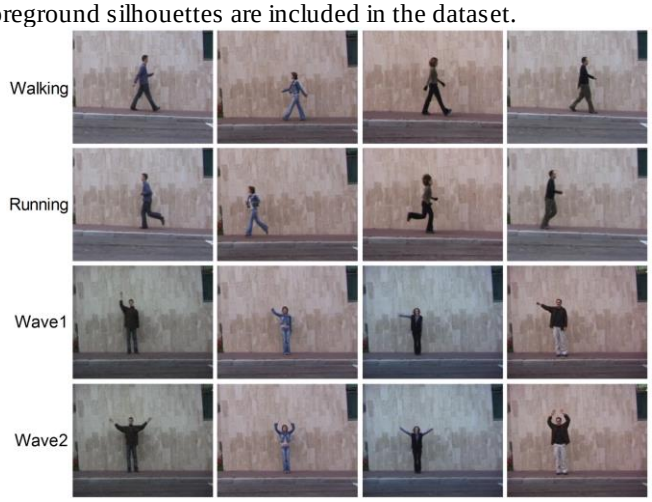

Fig. 4. Example frame for Weizmann Dataset

UCF Sport Action Dataset

Broadcast television UCF Sport action dataset consists of one fifty video sequences performed by thirteen actions type. Ten main actions such as dive, golf, kick, lift, ride, run, skateboard, swing band, swing side and walk. Most action classes there is considerable variation in action performance, human appearance, camera movement, view point, illumination and background.

IXMAS Dataset

The INRIA XMAS dataset contains eleven daily-life actions. Action such as chuckle watch, cross arms, scratch head, sit-down, get up, turn around, walk, wave, punch, kick, pick-up, performed each three times by eleven non professional actors. It contains four twenty nine multi view sequences. If the views are considered individually then it consists of 2145 sequences. The actions were filmed with five carefully calibrated and synchronized cameras.

Other Datasets

Datasets containing still images figure skating baseball and basketball are performed in [2]. The crowded videos dataset introduced in [3]. [4] Presented a set of still images collected from the web. The HOHA dataset [5] are a large collection of short segments of real Hollywood movies annoted with 12 action classes. The Hollywood human action data set contains eight actions extracted from movies and performed by a variety of actors. Stereo-based pedestrian detection in [6] benchmark dataset. Mono pedestrian detection study of pedestrian classification in [7] benchmark dataset. Occluded pedestrian classification benchmark dataset described in [8]. And many images or videos collected from the web.

Image Representation

The extracted features from the image sequences should generalize over the small variations in person appearance, view point, action execution and background. Image representation can be classified as two categories: Global representation and Local representation.

Global representation

In a top-down fashion the global representation is obtained. Initially, person localization in the image is analyzed by using background subtraction or tracking method and Region of interest (ROI) is prearranged which results in the image description. In general, Region of interest is obtained using background subtraction or tracking. The grid based approach will divide the observation into cells, each of cell encode part of the observation locally because global representation is derived from edges, silhouettes and optical flow and these are very sensitive to noise, partial occlusions and variations in viewpoints. Viewpoint insensitive action using envelope shape [9] focuses on use two orthogonally placed cameras at approximately similar height and similar to the person. Silhouettes from both cameras are aligned on the medial axis, and an envelope shape is calculated. The extract silhouettes from a single view and aggregate differences between subsequence frames of an action sequence are one of the earliest uses of silhouettes [10]. This result shows that the binary motion energy image (MEI) which indicates the motion occurs.

Region of Interest is divided into a fixed spatial or temporal grid small variation due to noise, partial occlusion and changes in viewpoint can be partly overcome by a global grid based representation and space time volumes. Each cell in the grid describes the image observation locally, and the matching function is changed accordingly from global to local [1]. These grids based representation resemble local representations, but require a global representation of the Region Of Interest. Optical flow in a grid-based representation is used by [11].

Space Time Volumes are often called as spatio-temporal volume (STV). By stacking frames over a given sequences 3 dimensional spatio-temporal volume is formed. To drive local space time saliency and oriented features the solutions of Poisson equation are used. In 3 dimensions spatio-temporal

volume background subtraction is not necessary, whereas 3-D super pixels are obtained from segmenting the STV. Global features for a given temporal range are obtained by calculating weighted moments over these local features. To construct an STV of flow and sample the horizontal and vertical components in space time [12] uses a 3-D variance of the rectangle features.

Local Representations

Local Representation contains a collection of local descriptor or patches to describe the observation. In this local representation the accurate localization and background subtraction are not essential. Local representation differs from the other representation; these are somewhat invariant to changes in viewpoint, person appearance and partial occlusion.

Local descriptors summarize an image or video patch in a representation that is ideally invariant to background clutter, appearance and occlusions, and possible to rotate and scale [1]. The spatial temporal size of a patch is usually determined by the scale of the interest point. Patches can also be described by local grid-based descriptors. Human action categories the frame or sequence can be represented as a bag-of-words, a histogram of codeword frequencies [14].

Local grid-based representation is similar to holistic approaches. Histograms of oriented gradients and flow are extracted at interest points, in a spatio-temporal grid. The position of the body and the size of the head is analyzed based on the grid spans. A subset of all possible blocks within the grid is selected using AdaBoost.

Space time interest point detectors locate the sudden changes of movement occur in the video. And these locations are informative for human action recognition. In space time interest point the local neighbor good has significant variation in both spatial temporal domains. Space time interest point detects subspace of correlated movement instead of detecting interest point over the entire volume. The difference between subsequence frames that estimate the focus of attention is calculated in [13].

Correlation between local descriptors approaches that exploit correlations between local descriptors for selection or the construction of high-level descriptions. Grid-based representation model temporal and spatial relations between local descriptors to some extent [1]. Learning deformable action templates introduce an active basis of shape and flow patches, where locations in space and time are allowed to very slightly [14]. Correlation between local descriptors can also be obtained by tracking features.

Image Segmentation

Images are considered as one of the most important mediums of conveying information in the field of computer vision. In computer vision segmentation refers to the process of partitioning a digital image into multiple segments (sets of pixels also known as super pixels).

Image segmentation for many years has been a high degree of attention. Algorithm development for one class of the image may not always be applied to another class of the image. There are many challenging issues like the development of a unified approach to image segmentation which can be applied to all types of images [15]. The different types of segmentations are available, such as pixel-based segmentation, edge based segmentation, region based segmentation, model based segmentation etc.,

Region based segmentation the image is portioned into connected regions by grouping neighboring pixels of similar intensity levels. In region growing [16] pixels in whole images are divided into sub regions or large regions based on predefined creations. Region splitting and merging is usually implemented with a theory based on quad tree data. The gradient is the first derivative for image f (x, y) when there is an abrupt change in intensity near edge and there is less image noise [17].

Thresholding operation converts a multiple image into a binary image based on a proper threshold. Segmentation based on clustering is an unsupervised learning task, where one needs to identify a finite set of categories known as clusters to classify pixels.

Background Subtraction

Background subtraction is a powerful mechanism for detecting changes in a sequence of image that finds many applications. Background subtraction is a process to segment the foreground object from the background of a video. Background subtraction is also an important component of many computer vision systems. There are two important steps in computer vision system one is to establish the background model and other background updates which separate the foreground and background.

Activity Detection

Human activity detection in videos is an important component of computer vision systems. The essential step is to identify the feature set that separates the human from the background even in cluttered scenes for identifying the activity performed.

Feature Extraction

The features are the useful information that can be extracted from the segmented human object by which the machine can understand the meaning of that posture. Feature extractions are the main vision task in action recognition and consist in extracting posture, hand gesture, facial expression, gait, behavior and motion cues from the video that are discriminating with respect to human action. The features are extracted from foreground images and it should be invariant to factors other than gait, human shape, such as texture, color or type of cloths. In most recognition approaches [18, 19], recognition features are extracted from silhouette images.

Spatio Temporal Features

The Spatio Temporal (ST) features have recently become a popular video representation for action recognition. The ST feature normally captures the strong variation of the data in spatial and temporal direction that are caused by motion of the actor. However, Spatio Temporal features contain only the appearance and motion information and ignore the shape of the information. Local space-time features capture characteristic shape and motion in video and provide relatively independent representation of events with respect to their spatio-temporal shifts and scales as well as background clutter and multiple motions in the scene. Such features are usually extracted directly from video and therefore avoid possible failures of other pre-processing methods such as motion segmentation and tracking [20].

Spin-Image Feature

Spin-images have been successfully used for object recognition. For actions, the Spin-images can provide a richer representation of how the local shape of the actor is changing with-respect to different reference points. These reference points may correspond to different limbs of the human body. Instead of attempting pairwise matching of Spin-images to match two actions, they use the bag of Spin-image strategy. First apply

PCA to compress the dimensionality of the spin-image, and then use K-means to quantize them. Then call the group of spin-images as a video-word. Finally, the action is represented by the bag of video-words model.

Mid-level Motion Feature

The Mid-level motion features focus on local regions of the image sequence. These features are tuned to discriminate between different classes of action, and are efficient to compute at run-time. Mid-level motion features are weighted combinations of threshold low-level features. Each mid-level feature covers small spatiotemporal cuboids, part of the whole figure-centric volume, from which its low-level features are chosen. Low level feature corresponds to a location in the figure-centric volume. For some small cuboids inside the figure-centric volume, using the Adaboost algorithm [22] to select a subset of the weak classifiers inside each figure-centric volume to construct better classifiers.

Low-Level Motion Features

To calculate the low-level motion features, first compute a figure centric spatio-temporal volume for each person. For each frame, low-level motion features are extracted from optical flow channels at the pixel locations in that frame and a temporal window of frames adjacent to it. These low-level features individual locations are not capable of discriminating between the positive and negative classes e.g. two different action categories) much better than random classification.

Skeletal Feature

The Skeletal features extraction and separating human body model into several human body parts like face, torso, hand and limbs. Human action recognition system based on segmented skeletal features which are separated into several human body parts. Skeleton based object recognition systems generally perform better than shape based object recognition approaches [23]. This work extracts and split the human skeleton using Normalized Gradient Vector Flow in the space of diffusion tensor fields, using the eigen values and eigen vectors of the segmented skeletal features. To extract the robust features of a target object to effectively understand human behaviors form Skeletal feature.

Contour-Based Feature

Contour-based feature representations have a long history in object recognition and computer vision. In contour-based approaches, often the first step is detected from edges. To extract the contours, the document images first need to binaries. In this method, instead of tracking the whole set of pixels comprising an object, the algorithm tracks only the contour of the object. In [24] proposed contour based nonridge object tracking method via the contour energy function. Tracked the complete region of the nonridge objects and recovered the occluded object parts.

Activity Recognition

Recognition or classification of hand gestures is the last phase of the recognition system. Hand gestures can be classified using two approaches. These approaches are Rule based approaches and Machine learning based approaches.

Rule Based Approach

Rule based approaches are represented the input features as a manually encoded rule, and the winner gesture is the one that matched with the encoded rules after his features has been extracted. The main problem with this technique is that the human ability is encoding the rule's limits the success of the recognition process [25].

Machine learning based approaches

Machine learning based algorithm can be divided into supervised and unsupervised approaches.

Recognition using the HMM

In Markov Model (MM) the state is directly visible to the observer, so the state transition probability is the only parameter. The Hidden Markov Model (HMM) models are sequences of observations as a piecewise stationary process. In hidden Markov model the states are not directly accessible to the observer. Each state has probability distribution over output tokens. The sequence of tokens generated by an HMM gives some information about the sequence of statements. Hidden variable controls the components to be selected for each observation. The HMM is stochastic approach which models the given problem as a ”doubly stochastic process” in which the observed data are thought to be the result of having passed the Hidden processes are to be characterized using only the one that could be observed. The HMM will be useful in real world applications, if the following three basic problems of HMM are solved [26] Such as Evaluation problem, Decoding problem, Learning problem.

The following is an illustrative list of applications of HMM:

Speech recognition
Gait recognition
Optical character recognition
Lip-reading (visual speech to text mapping)
Gesture and body motion analysis

Two types of HMM model such as ergodic model and Left-Right model. The Hidden Markov Models are a popular technique for recognizing human gesture in a variety of applications and sensor configuration. Hidden Markov Models are double stochastic process as governed by an underlying Markov chain with a finite number of states, and a set of random functions each of which is associated with one state [29]. Several hand gesture recognition systems have been developed using various features computed from static images or image sequences [30]. The vision-based method selects the input data as the feature vectors for the HMM input and other HMM-based [31, 32] hand gesture recognition systems have also been developed. Gesture recognition systems using HMM models has been developed [33].

Recognition using the SVM

Support Vector Machine (SVM) is a kernel-based technique which is based on the principle of structured risk minimization (SRM). SVM constructs a linear class boundary based on support vectors. SVM are sets of related supervised learning model with an associated learning algorithm that analyze data and recognize patterns used for classification and recognition. SVMs [34] perform pattern classification between two classes by finding a decision surface that has a maximum distance from the closest points in the training set.

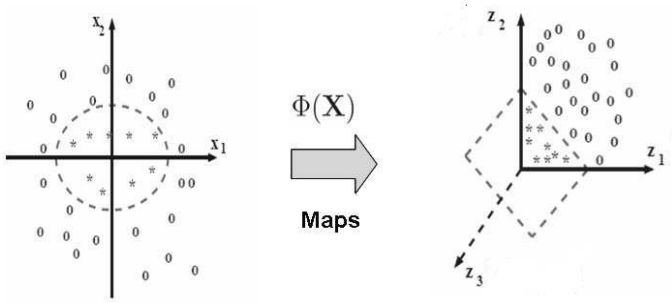

Fig. 5. Non-leaner, Linear structure

Vision based gesture recognition for alphabetical hand gesture recognition using SVM [35]. This gesture recognition system for alphabetical hand gesture is built. The system is designed using the Support Vector Machines Classifier which is widely used for classification and regression testing.

Recognition using the KNN

K-nearest neighbors (KNN) classifiers have a good performance when the attributes of a system are linearly separable. The class which has the most vectors in those K neighbors is chosen to be the class of the input vector. A cluster is a collection of objects which are similar between them and are dissimilar to the objects belonging to other clusters. Clustering is an unsupervised learning method which deals with finding a structure in a collection of unlabeled data. A loose definition of clustering could be the process of organizing objects into groups whose members are similar in some way.

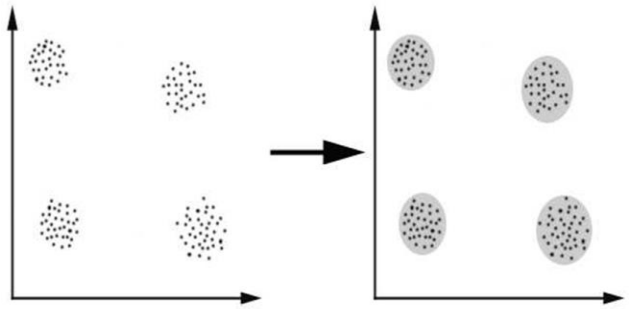

Fig. 6. Clustering

K-means clustering [26, 28] is an algorithm to group objects based on attributes/features into k number of groups where k is a positive integer. The rouping (clustering) is done by minimizing the Euclidean distance between the data and the corresponding cluster centroid. Thus the purpose of k-means clustering is to cluster the data. They calculate the distance between the cluster centroid to each object using the Euclidean distance measure. E-M algorithm [26, 28] finds out maximum likelihood estimates of parameters in probabilistic models. This algorithm iterates between the E- step and the M step until convergence. Expectation step computes an expectation of the likelihood assuming parameters. Maximization step computes maximum likelihood estimates of parameters by maximizing the expected likelihood found in the E - step.

K-nearest Neighbors with Distance Weighting (KNNDW) is an improvement which has been proved to perform better than KNN in many cases [36]. In this method, the contribution of each neighbor to the overall classification is weighted by its distance from the point being classified.

Recognition using the NN

Artificial Neural Network (ANN) is an information processing system that is inspired by biological nervous systems like the brain process information. A neural network is a machine that is designed to model the way in which the rain performs a particular task or functions; the network is usually implemented by using electronic components or simulated in software on a digital computer. To achieve good performance, neural networks employ massive interconnections of simple computing cells referred to as “neurons” or “processing units”. Gesture recognition is an important for developing alternative human-computer interaction modalities using ANN.

Types of Activation function such as Linear/Threshold function, Sigmoid/Squashing function and hyperbolic tangent function. Artificial Neural networks are flexible in a changing environment [37] and It also describes the process of gesture recognition using ANN [38] Continuous time Recurrent Neural Network (RNN) is used. This approach is based on the idea of creating specialized signal predictors for each gesture class.

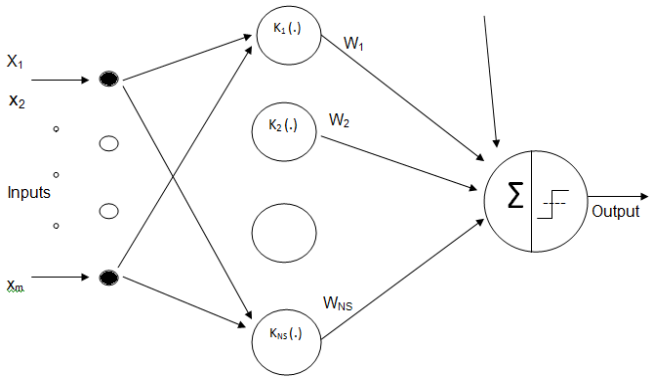

Fig. 7. Structure of artificial neuron

Application Areas

Gestures are modeled as sequences of multiple events. Every event is matched independently with its own event model and linear time scaling. Gesture recognition constitutes matching the appropriate events sequentially. Recognition of gestures is performed using a probabilistic finite state machine. A human hand has abundant joints.

It is able to form several distinct hand shapes. Hand shapes are quite useful to represent different communication signs, which are defined to execute specific tasks.

Sign Language

In vision based gesture recognition, hand shape segmentation is one of the toughest problems under a dynamic environment. It can be simplified by using visual marking on the hands. Some researchers have implemented sign language and pointing gesture recognition based on different marking modes [39]. The Hidden Markov Model has basic properties that make them very attractive for dynamic gesture recognition [40] that develops a gesture recognition system for sign language recognition.

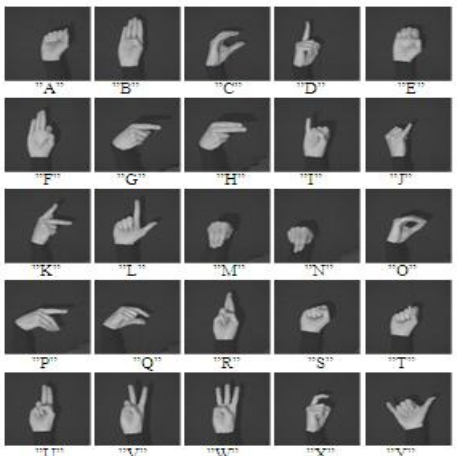

Fig. 8. American sign language

Robotics

Robotics, human manipulation and interaction researches used postures and gestures to learn the robot some interaction commands by explaining its appropriate meaning for the robot as an action [41]. Gestures are used for controlling robots, corresponding to the virtual reality interaction system. For virtual reality application gestures are considered as one of the effective spreading stages in the computing area [42].

Fig. 9. Human Robot interaction taken from [43]

Gesture-to-speech application converts hand gestures into speech. This system enables hearing-impaired people to communicate with their surrounding environments through computers and interacts easily with other people even without knowing the sign language [41].

Implemented computer vision and gesture recognition techniques developed a vision based low cost input device for controlling the VLC player through gestures [44]. Computer games applied gesture recognition on virtual game application [45] and hierarchical recognition of international human gestures for sports.

A television control system developed by hand gestures using an open hand and the user can change the channel, turn the television off and on, decrease, increase and mute the volume [46].

Traffic police hand signal recognition

The traffic police gesture systems are mainly expressed by arms. According to Indian traffic rules there are 12 hand gestures. The Chinese traffic police gesture system is defined and regulated by the Chinese ministry of public security [47] Gesture of traffic police officers are captured in the form of depth images. In [48] Road traffic control system it considers only the arm directions for classifying the traffic control commands. Traffic gesture using three types of control commands like six defined traffic gestures are used.

Fig. 10. General Rules for traffic police

Challenges and Future Work

Besides the computer vision issues, an action recognition problem has to face a number of contextual and philosophical challenges. A system must recognize, whether a person is running to catch a bus or running out of committing a crime.

Challenges

The major challenges in human action recognition are listed here the common approach is to extract image features of the video and to issue a corresponding action class label. Several actions, there are large variations in performance such as Intra and Inter-class variations. For example, running movements can be different in speed and stride length. A good human action recognition approach should be able to generalize over variations within one class and distinguish between actions of different classes. For increasing the numbers of action classes, this will be more challenging as the overlap between classes will be higher. The environment in which the human action performance takes place is an important source of variation in the recording. The person of the party might be occluded in the recording. Human localization might prove harder in clutter or dynamic environments. Lighting condition can further influence the appearance of the person. The same action, observed from different viewpoints, can lead to very different image observations. Dynamic backgrounds increase the complexity of localizing the person in the image and robustly observing the motion. When using a moving camera, these challenges become even harder. In vision-based human action recognition, all these issues should be addressed explicitly.

Future Direction

Most of the work reported in the literature is restricted to the single viewpoint. Applying multiple viewpoint is needed to solve this issue but with added training complexity. Further, action recognition in traffic surveillance requires real-time processing. And, common datasets focus on particular application domain.

Motivated by the wide range of applications, this paper represented the current state of the art research in action recognition. If the challenges ahead of this research are fulfilled, this would be a great step towards achieving a robust interpretation and recognition of actions in the near future.

Conclusion

This paper discusses about the detailed overview and categories of current issues and trends. Various applications of human action recognition are discussed in this paper. Common datasets like KTH dataset, Weizmann dataset, UCF dataset, IXMAS dataset and other datasets are discussed. In image representation, global image representation can be extracted with low cost and it also achieves good results. In this paper various feature extraction methods in activity detection are discussed.

In this survey basic concepts and techniques behind the gesture recognition system has been studied. Vision based approaches are a widely used method in gesture recognition since, it is a very realistic approach and also it provides better results while comparisons to the data glove approaches. The study of the hand gesture recognition and its various applications were discussed in this paper. The hand gesture of a person is analyzed by segmenting the object from the environment. So, objects in the environment have to be segmented by the feature extraction. The different types of techniques to recognize the hand gesture are reviewed and analyzed.

Human machine interaction can be achieved by different recognition technique and the gesture can be recognized even if they are not performed perfectly.

References

[1] DoHyung Kim, Jaeyeon Lee, Ho-Sub Yoon Jaehong Kim, Joochan Sohn. Vision-based arm gesture recognition for a long-range human robot interaction. Electronics and

Telecommunications Research Institute, Daejeon, Korea. 2013 July; 65 (1): 336-352.
[2] Yang Wang, Hao Jiang, Mark S, Drew, Ze-Nian Li and Greg Mori. Unsupervised discovery of action classes. In: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR06). 2006 June; 2: 1654-1661.
[3] Yan Ke, Rahul Sukthankar and Martial Hebert. Event detection in crowded videos. In: Proceedings of the International Conference On Computer Vision (ICCV07). Rio de Janeiro, Brazil. 2007 October; 18.
[4] Nazli Ikizler, Ramazan G. Cinbis, Selen Pehlivan and Pinar Duygulu. Recognizing actions from still images. In: Proceedings of the International Conference on Pattern Recognition (ICPR08). Tampa, FL. 2008 December; 14.
[5] Laptev I, Marszalek M, Schmid C, Rozenfeld B. Learning realistic human actions from movies. In: Conference on Computer Vision and Pattern Recognition. 2008; 18.
[6] Keller C, Enzweiler M, and Gavrila D.M. A New Benchmark for Stereo-based Pedestrian Detection. Proc. Of the IEEE Intelligent Vehicles Symposium. 2011 June 5-9; 691-696.
[7] Munder S, Gavrila D.M. An Experimental Study on Pedestrian Classification IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006 November; 28: 1863-1868.
[8] Enzweiler M, Eigenstetter A, Schiele B and Gavrila D.M. Multi-Cue Pedestrian Classification with Partial Occlusion Handling. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2010.
[9] Feiyue Huang, Guangyou Xu. Viewpoint insensitive action recognition using envelope shape. In: Proceedings of the Asian Conference on Computer Vision (ACCV07) part 2, Lecture Notes in Computer Science. Tokyo, Japan. 2007 November; 477-486.
[10] Aaron F, Bobick, James, Davis W. The recognition of human movement using temporal templates. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI). 2001 March; 23 (3): 257-267.
[11] Somayeh Danafar, Niloofar Gheissari. Action recognition for surveillance applications using optical flow and SVM. In: Proceedings of the Asian Conference on Computer Vision (ACCV07) part. Lecture Notes in Computer Science, Tokyo, Japan. 2007 November; 4844: 457-466.
[12] Paul A, Viola, Michael, Jones J. Rapid object detection using a boosted cascade of simple features. In: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR01). 2001 December; 1: 511-518.
[13] Matteo Bregonzio, Shaogang Gong, Tao XiangRecognising action as clouds of space-time interest point. In: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR09). Miami, FL. 2009 June; 1-8.
[14] Juan Carlos Niebles, Hongcheng Wang, Li Fei-Fei. Unsupervised learning of human action categories using spatial temporal words. International Journal of Computer Vision (IJCV). 2008; 79 (3): 299-318.
[15] Duda R. O, Hart P. E. Pattern Classification and Scene Analysis. New York: Wiley. 1973.
[16] Matteo Bregonzio, Shaogang Gong and Tao Xiang. Recognizing action as clouds of pace time interest points. In: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR09). Miami, FL. 2009 June; 1-8.
[17] Ahmad, Seong-Whan Lee. Human action recognition using shape and CLG-motion flow of multi-view image sequences. Pattern Recognition. 2008; 41 (7): 2237-2252.
[18] Freeman W. T and Weissman C. D. Television control by hand gestures. IEEE International Workshop on Automatic Face and Gesture Recognition. 1995.
[19] Phillips P.J, Sarkar S. Robledo I, Grother P and Bowyer K. The gait identification challenge problem: Data sets and baseline algorithm. In International Conference on Pattern Recognition. 2002; 1: 385 – 388.
[20] Laptev I. Marszalek M. Schmid c and Rozenfeld b. Learning realistic human actions from movies, In CVPR. 2008 June 23-28; 1-8.
[21] Jingen Liu, Saad Ali, Mubarak Shah. Recognizing human actions using multiple features. IEEE Conference on Computer Vision and Pattern Recognition. 2008 June 23 - 28; 1-8: 1063-6919.
[22] Viola P and Jones M. Robust real-time object detection. In 2nd Intl. Workshop on Statistical and Computational Theories of Vision. 2001 Febuary; 2: 3, 4.
[23] Sang Min Yoon, Kuijper A. Human Action Recognition using Segmented Skeletal Features. 20th International Conference on Pattern Recognition (ICPR). 2010 august; 3740–3743.
[24] Yilmaz A. Li, X. Shah M. Contour-based object tracking with occlusion handling in video acquired using mobile cameras. IEEE Trans. Pattern Anal. Mach. Intell. 2004;26(11): 1531-1536.
[25] Murakami K and Taguchi H. Gesture recognition using recurrent neural networks. ACM, Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technological. 1999; 91: 237-242.
[26] Duda R.O. Hart P.E and Stork D.G. Pattern Classification, John Wiley and Sons, Singapore. 2003.
[27] Lawrence R Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. 1989 Feb; 77 (2): 257-286.
[28] Haykin S. Neural networks: A comprehensive foundation. Prentice Hall International, New Jersey. 1999.
[29] Oren Boiman, Michal Irani. Detecting irregularities in images and in video. International Journal of Computer Vision (IJCV). 2007; 74 (1): 17-31.
[30] Rabiner L.R. A tutorial on hidden Markov models and selected application in speech recognition, Proc. IEEE 77. 1989; 267–293.
[31] Huang T.S. Pentland A. Hand gesture modeling. Analysis, and synthesis. Proceedings of the International Workshop on Automatic Face-and Gesture-Recognition, Zurich, Switzerland. 1995 June; 73–79.
[32] Campbell L.W, Becker D. A, Azarbayejani A, Bobick A.F, Plentland A. Invariant features for 3-D gesture recognition. Proceedings IEEE Second International Workshop on Automatic Face and Gesture Recognition. 1996 October 14-16; 157-162.
[33] Campbell L.W, Becker D. A, Azarbayejani A, Bobick A.F, Plentland A. Invariant features for 3-D gesture recognition, Proceedings IEEE Second International Workshop on Automatic Face and Gesture Recognition. 1996.
[34] Mahmoud Elmezain, Ayoub Al-Hamadi, Jorg Appenrodt and Bernd Michaelis. A Hidden Markov Model-Based Isolated and Meaningful Hand Gesture Recognition. International Journal of Electrical and Electronics Engineering. 2009; 156–163.
[35] Cortes C and Vapnik V. Support vector networks, Machine learning. 1995; 20: 53–60.

[36] Aseema Sultana, T Rajapuspha. Vision Based Gesture Recognition for Alphabetical Hand Gestures Using the SVM Classifier. Aseema Sultana et al. International Journal Of Computer Science and Engineering Technology (IJCSET). 2012 July; 3: 7.
[37] Pujan Ziaie, Thomas Miller, Mary Ellen Foster, Alois Knoll. Using a Nave Bayes Classifier based on K Nearest Neighbors with Distance Weighting for Static Hand- Gesture Recognition in a Human-Robot Dialog System. 2007 November.
[38] Miss. Shwetah K. Yewale MR, Pankaj K, Bharne. Artificial neural network approach for hand gesture recognition. Shweta, K. You-all et al. International Journal of Engineering Science and Technology (IJEST). 2011 April; 3,4.
[39] Fan Guo, Zixing CAI, Jin Tang. Chinese Traffic Police Gesture Recognition in Complex Scene. International Joint Conference of IEEE TrustCom-11/IEEE ICESS-11/FCST-11. 2010.
[40] James Davis and Mubarak Shah. Recognizing hand gestures. ECCV, Stockholm, Sweden. 1994 May; 331–340.
[41] Starner t. Visual Recognition of American Sign Language Using Hidden Markov Models. Master's Thesis. MIT. 1995 Febuary.
[42] Murakami K and Taguchi H. Gesture recognition using recurrent neural networks. Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technological, ACM. 1999; 237–242.
[43] Murthy G. R. S and Jadon R. S. A review of vision based hand gestures recognition International Journal of Information Technology and Knowledge Management. 2009; 2(2); 405–410.
[44] Mitra S and Acharya T. Gesture recognition: A survey. IEEE Transactions on systems. Man and Cybernetics. Part C: Applications and review. 2007; 37(3): 311–324.
[45]] Rautaray S. S and Agrawal A. A vision based hand gesture interface for controlling VLC media player. International Journal of Computer Applications. 2010; 10(7):
[46] Starner T. Weaver J and Pentland A. Real-time American Sign Language recognition using a desk and wearable computer based video. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002; 20(12): 1371–1375.
[47] Pavlovic V. I, Sharma R and Huang T. s. Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review, IEEE Transactions on Pattern Analysis and Machine Intelligence. 1997; 19(7): 677–695.
[48] Ben Wang and Tao Yuan. Traffic Police Gesture Recognition using Accelerometers. 2008; 1: s4244–2581.